A uniform analysis of HD 209458b *Spitzer*/IRAC lightcurves with Gaussian process models

Thomas M. Evans,[1,2*] Suzanne Aigrain,[1] Neale Gibson,[3] Joanna K. Barstow,[1] David S. Amundsen,[2] Pascal Tremblin[2,4] and Pierre Mourier[2,5]

[1] *Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[2] *School of Physics, University of Exeter, EX4 4QL Exeter, UK*
[3] *European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany*
[4] *Maison de la Simulation, CEA-CNRS-INRIA-UPS-UVSQ, USR 3441, Centre d'étude de Saclay, 91191 Gif-Sur-Yvette, France*
[5] *Master ICFP, Département de Physique, Ecole Normale Supérieure, 24 Rue Lhomond, 75005 Paris, France*

ABSTRACT

We present an analysis of *Spitzer*/IRAC primary transit and secondary eclipse lightcurves measured for HD 209458b, using Gaussian process models to marginalise over the intrapixel sensitivity variations in the 3.6μm and 4.5μm channels and the ramp effect in the 5.8μm and 8.0μm channels. The main advantage of this approach is that we can account for a broad range of degeneracies between the planet signal and systematics without actually having to specify a deterministic functional form for the latter. Our results do not confirm a previous claim of water absorption in transmission. Instead, our results are more consistent with a featureless transmission spectrum, possibly due to a cloud deck obscuring molecular absorption bands. For the emission data, our values are not consistent with the thermal inversion in the dayside atmosphere that was originally inferred from these data. Instead, we agree with another re-analysis of these same data, which concluded a non-inverted atmosphere provides a better fit. We find that a solar-abundance clear-atmosphere model without a thermal inversion underpredicts the measured emission in the 4.5μm channel, which may suggest the atmosphere is depleted in carbon monoxide. An acceptable fit to the emission data can be achieved by assuming that the planet radiates as an isothermal blackbody with a temperature of 1484 ± 18 K.

Key words: planets and satellites: atmospheres; planets and satellites: general; methods: data analysis, astronomical instrumentation, methods, and techniques; stars: individual: HD209458

1 INTRODUCTION

Over the past decade, the *Spitzer Space Telescope* has proven to be a productive facility for characterising the atmospheres of transiting exoplanets (e.g. Charbonneau et al. 2005; Deming et al. 2006; Knutson et al. 2007a; Désert et al. 2009; Crossfield et al. 2012; Lewis et al. 2013; Todorov et al. 2014). The ability of its instruments to probe the $\sim$3–25μm wavelength range has provided constraints on the thermal emission of numerous exoplanets, as well as atmospheric transmission in a region dominated by absorption from molecular species such as water, methane, carbon monoxide, and carbon dioxide. Hot Jupiters have offered especially favourable targets for such observations, given their large atmospheric

scale heights and relatively strong emission at these wavelengths.

This paper focuses on observations made with the Infrared Array Camera (IRAC), which has been the most widely used *Spitzer* instrument for observing exoplanets. Specifically, we analyse ten transits and eleven eclipses that have been measured for HD 209458b, most of which have already been published (Knutson et al. 2008; Beaulieu et al. 2010; Zellem et al. 2014; Diamond-Lowe et al. 2014). By providing a uniform analysis of these datasets, we aim to re-evaluate a number of claims that have been made in the literature. In particular, Beaulieu et al. (2010) measured significantly larger effective radii for the planet in the 5.8μm and 8.0μm channels relative to the 3.6μm and 4.5μm channels in transmission, and interpreted this as evidence for water absorption. However, Deming et al. (2013) have since resolved the water absorption band centred at 1.4μm us-

* E-mail: tevans@astro.ex.ac.uk

ing the *Hubble Space Telescope* (HST) Wide Field Camera 3 (WFC3), and found that it has a much lower amplitude than would be expected based on the results of Beaulieu et al. We also seek to address the claim of a thermal inversion in the dayside atmosphere, which was first postulated by Knutson et al. (2008) based on the deeper eclipses those authors measured for the 4.5μm and 5.8μm channels relative to the 3.6μm channel. The former channels coincide with absorption features due to water and carbon monoxide: therefore, seeing these features in emission would suggest an increasing temperature profile with decreasing pressure. Diamond-Lowe et al. (2014) have challenged this picture by presenting revised eclipse depths that are suggestive of a non-inverted pressure-temperature profile. Furthermore, Hansen et al. (2014) have suggested that the emission data for HD 209458b are consistent with radiation from an isothermal blackbody.

The conflicting results obtained by different authors analysing the same datasets is likely due to the various methods that have been used to account for the instrumental systematics that dominate IRAC lightcurves. The main contribution of the current study is to apply the machinery of Gaussian processes (GPs) to the task of treating these systematics. This work follows similar applications of GP models to transit lightcurves published by Gibson et al. (2012b,a, 2013a,b), Evans et al. (2013), and Gibson (2014). One of the primary advantages of GP models is that they allow us to naturally handle correlations in the data that may be poorly understood from a first principles standpoint, by specifying only high-level properties of the covariance. This relaxes the assumptions built into our model, by removing the need to associate the systematics with a deterministic functional form. The resulting model is less restrictive, allowing us to capture a broad range of systematics behaviours with a relatively small number of tunable parameters. Furthermore, GPs are Bayesian models in the sense that uncertainty is treated transparently using self-consistent rules of probability. Each unknown in our model is associated with a probability distribution that reflects our uncertainty in its value, and it is possible to write down an expression for the likelihood of the observed data given specific values for the model parameters, i.e. the model posterior distribution. We can then optimise the model posterior with respect to the unknown parameters, or marginalise over the parameter space using a method such as Markov chain Monte Carlo (MCMC). These properties make GP models suitable for inferring planet parameters from transit lightcurves affected by systematics that are not especially well-understood, such as those obtained with IRAC.

The paper is arranged as follows. Section 2 describes the lightcurve observations analysed for this study and Section 3 describes how we produced lightcurves from the raw data frames. IRAC instrumental systematics are described in Section 4, with an overview of methods that have been used to correct for them previously in the literature. Section 5 outlines the GP methodology that we adopt in the current study, and Section 6 describes the lightcurve fitting. The results are presented in Section 7 and discussed in Section 8, with a focus on the implications for the planet atmosphere. Our conclusions are summarised in Section 9.

2   OBSERVATIONS

We have analysed ten primary transits and eleven secondary eclipses for HD209458b, made across all four IRAC channels. Details of the relevant *Spitzer* observing programs are given in Table 1, along with references to previously published analyses. More specific information for the individual lightcurves is given in Table 2. For this study, we did not model the complete half- and full-phase datasets acquired for Programs 30825, 40280, and 60021. Instead, for these lightcurves only ~ 5 hr subsections centred on the transits and eclipses were analysed.

Most observations were made in stare mode. The only exceptions were those made as part of Program 20523, which have been published in Knutson et al. (2008). For the latter, four sets of 64 frames were acquired in a given channel before the telescope was repointed to be centred on the next channel, and another four sets of 64 frames were taken. This process was cycled through each of the four channels, and repeated for the duration of the observations. Knutson et al. discarded the first set of 64 frames in each set of four, as the star was still drifting significantly during this time following the repointing. In addition, those authors discarded the first ten frames and the 58th frame from each set of 64 for the 5.8μm and 8.0μm channels, as these exhibited count levels consistently below the median. We performed two separate analyses for these lightcurves: one with this culling applied, and one without. However, we obtained consistent results in both cases, and only present results for the unculled dataset below.

3   DATA REDUCTION

The Basic Calibrated Data (BCD) frames for each lightcurve were reduced using a custom pipeline written in the Python programming language.[1] The first step performed by the pipeline is to calculate the background level and locate the stellar centroid in each BCD frame. The background was estimated by taking the median pixel value from the four 8×8 pixel subarrays at the corners of each frame, and then subtracted from each pixel in the array. The centroid coordinates were then determined by taking the flux-weighted mean of a 7×7 pixel subarray centred on the approximate location of the star. An initial guess was provided for the stellar centroid coordinates in the first frame, and coordinates determined for the previous frame were used as the initial guess in subsequent frames. Mid-times were computed for each exposure in Barycentric Julian Date Coordinated Universal Time (BJD$_{\rm UTC}$) using the BMJD_OBS and FRAMTIME header entries. Bad frames were flagged by identifying those with outlying centroid coordinates or pixel counts. This was done by comparing against the median and standard deviation of the 30 frames immediately preceding and following each frame. If the centroid coordinates or any pixel counts within a subarray spanning the photometric aperture centred at the stellar centroid differed from the median by $> 5\sigma$, the frame was discarded from the analysis. This process was iterated twice, resulting in 0.2–4.4% of the frames being discarded depending on the dataset (Table 2).

---

[1] Publicly available at www.github.com/tomevans

Table 1. HD209458 datasets analysed for this study.

Program	P.I.	Type	Channels (μm)				References[a]
			3.6	4.5	5.8	8.0	
20523	D. Charbonneau	Eclipse	Yes	Yes	Yes	Yes	Kn08, DL14
40280	H. Knutson	Half-phase	-	-	-	Yes	DL14
461	G. Tinetti	Transit	Yes	Yes	Yes	Yes	Be10
60021	H. Knutson	Full-phase	Yes	Yes	-	-	Ze14, DL14

[a] Be10 (Beaulieu et al. 2010), DL14 (Diamond-Lowe et al. 2014), Kn08 (Knutson et al. 2008), Ze14 (Zellem et al. 2014).

Table 2. Lightcurve details.

Program	Type	Channel (μm)	Date (UT)	Mode[a]	Flagged[b] (%)	N_b[c]	η_b[d]	Δt_{med}[e] (sec)	r_{ap}[f] (pix)
20523	Eclipse	3.6	2005 Nov 28	sub, 0.1	0.36	1115	32	4.5	4.0
	Eclipse	4.5	2005 Nov 28	sub, 0.1	0.21	1115	32	4.5	3.0
	Eclipse	5.8	2005 Nov 28	sub, 0.1	0.32	1115	32	4.5	2.5
	Eclipse	8.0	2005 Nov 28	sub, 0.1	0.55	1115	32	4.5	3.0
40280	Transit	8.0	2007 Dec 25	sub, 0.4	1.11	1577	32	13.5	3.5
	Eclipse	8.0	2007 Dec 24	sub, 0.4	1.12	1577	32	13.5	4.0
461	Transit	3.6	2007 Dec 31	full, 0.4	1.68	1427	2	16.8	2.5
	Transit	3.6	2008 Jul 19	full, 0.4	2.03	1428	2	16.8	2.5
	Transit	4.5	2008 Jul 22	full, 0.4	1.67	1288	2	16.8	2.5
	Transit	5.8	2007 Dec 31	full, 2.0	3.78	1425	2	16.8	3.5
	Transit	5.8	2008 Jul 19	full, 2.0	3.81	1425	2	16.8	3.0
	Transit	8.0	2008 Jul 22	full, 2.0	4.42	1286	2	16.8	4.0
60021	Eclipse (1st)	3.6	2011 Jan 12	sub, 0.1	0.25	1285	128	16.8	2.5
	Transit	3.6	2011 Jan 14	sub, 0.1	0.25	1286	128	16.8	2.5
	Eclipse (2nd)	3.6	2011 Jan 16	sub, 0.1	0.29	1286	128	16.8	3.0
	Eclipse (1st)	3.6	2014 Feb 13	sub, 0.1	0.23	1286	128	16.8	3.0
	Transit	3.6	2014 Feb 15	sub, 0.1	0.23	1286	128	16.8	3.0
	Eclipse (2nd)	3.6	2014 Feb 17	sub, 0.1	0.18	1285	128	16.8	2.5
	Eclipse (1st)	4.5	2010 Jan 18	sub, 0.4	0.62	1577	32	13.6	2.5
	Transit	4.5	2010 Jan 19	sub, 0.4	0.73	1575	32	13.6	3.0
	Eclipse (2nd)	4.5	2010 Jan 21	sub, 0.4	0.62	1577	32	13.6	2.5

[a] Readout mode and frame time in seconds.
[b] Fraction of frames flagged as bad.
[c] Number of frames in the binned dataset used for the GP lightcurve fit.
[d] Number of consecutive frames per bin.
[e] Median cadence of binned frames.
[f] Photometric aperture radius.

Photometry was performed for each remaining frame by summing the pixel counts within circular apertures. Separate reductions were obtained for different aperture radii, ranging between 2–6 pixel in increments of 0.5 pixel. Due to the undersampled nature of the IRAC point spread func-tion (PSF), we linearly interpolated the native pixel array onto a 10×10 super-sampled grid, as has been done by others previously (e.g. Stevenson et al. 2010). These inter-polated sub-pixels were counted towards the aperture sum

if their centres fell within the aperture radius. The resulting lightcurves are shown in Figure 1.

4   INSTRUMENTAL SYSTEMATICS

The raw lightcurves are affected by instrumental systematics that are characteristic of IRAC and have been documented extensively in the literature (e.g. Charbonneau et al. 2005; Agol et al. 2010; Seager & Deming 2010; Stevenson et al. 2012). The systematics divide into two broad categories: intrapixel sensitivity variations in the 3.6μm and 4.5μm channels, and the ramp effect in the 5.8μm and 8.0μm channels.

4.1 3.6μm and 4.5μm channels

In the 3.6μm and 4.5μm channels, which employ InSb detectors, the measured flux correlates with the position of the stellar PSF on the detector array. As noted in the IRAC instrument handbook,[2] this effect is believed to be caused by variations in the quantum efficiency across individual pixels. Pointing drift during observations, combined with the undersampled PSF, therefore results in variations in the measured flux at the level of a few percent.

Traditionally, intrapixel sensitivity variations have been treated in IRAC data by decorrelating the lightcurve against a low-order polynomial in the xy centroid coordinates of the stellar PSF, which can either be removed prior to fitting the planet signal or fit simultaneously with the planet signal (e.g. Charbonneau et al. 2008; Knutson et al. 2008; Désert et al. 2009). The main issue with this method is that a low-order polynomial may not have the flexibility to fully capture the underlying correlations in the data if there is fine-scale structure present. This could be addressed to some extent by continuing to add higher order terms to the polynomial decorrelation; however, such an approach runs the risk of overfitting and increases the dimensionality of the parameter space that must be marginalised over using a method such as MCMC.

An alternative decorrelation method has been suggested by Ballard et al. (2010), which uses a 2D Gaussian convolution to construct a smoothed pixel sensitivity map from the flux measurements. Using this approach, the authors identified a high-frequency corrugation structure in the xy sensitivity for a 4.5μm lightcurve that would not have been captured by a low-order polynomial decorrelation. Along similar lines, Stevenson et al. (2012) have proposed the use of bilinear interpolation of the measured fluxes to map intrapixel sensitivity variations, which is sensitive to spatial scales corresponding to the grid of knots used. Knutson et al. (2012) and Lewis et al. (2013) have also identified a correlation between the noise pixel parameter β of a given data frame – which is inversely proportional to the PSF sharpness parameter, described in Muller & Buffington (1974) – and the measured flux. This can be treated by either using variable photometric apertures for each frame that scale with β, or including β as a decorrelation variable in the systematics treatment. More recently, Deming et al. (2014) have presented a pixel-level decorrelation (PLD) method, which

---

[2] http://irsa.ipac.caltech.edu/data/SPITZER/docs/irac

works by linearly decorrelating the measured fluxes against the individual pixel counts within a subarray centred on the PSF, rather than using the centroid coordinates directly.

4.2 5.8μm and 8.0μm channels

In the 5.8μm and 8.0μm channels, which employ Si:As detectors, the measured flux smoothly increases or decreases before levelling off. As with the intrapixel sensitivity variations, the amplitude of this effect is usually a few percent, with the steepest change in flux occurring during the initial $\sim 1\,$hr of observations. The ramp can be attributed to the sensitivity of individual pixels varying as a function of time, where the rate of change depends on the illumination level of the pixel (e.g. Knutson et al. 2007a).

It has been suggested that the ramp is caused by electrons getting caught in charge traps within the detector (e.g. Deming et al. 2007; Agol et al. 2010; Seager & Deming 2010). As more photons arrive, the charge traps fill up, resulting in less electrons getting trapped and a higher flux being measured. However, Seager & Deming (2010) acknowledge that the physics of the detectors are poorly understood and this simple picture may be incomplete or wrong. For instance, lightcurves measured in the 5.8μm channel exhibit a ramp-down behaviour (Figure 1), which is not obviously explained by charge trapping.

Standard practise has been to fit 5.8μm and 8.0μm lightcurves by multiplying the transit signal by a parametric model that provides a good approximation to the ramp. One approach is to remove the initial steep section of the lightcurve and fit the remainder with a linear or quadratic polynomial in time (e.g. Deming et al. 2007; Beaulieu et al. 2010). However, a more common approach is to model the full baseline with a low-order polynomial in logarithmic time (e.g. Charbonneau et al. 2008; Knutson et al. 2008; Désert et al. 2009; Machalek et al. 2010) or exponential time (e.g. Agol et al. 2010).

Subtle differences between the various ramp functions used to model 5.8μm and 8.0μm lightcurves can potentially bias the inferred planet parameters. For example, polynomials in exponential time tend to retain slightly more curvature than polynomials in logarithmic time after the initial steep gradient. This can result in an underestimated transit depth if the ramp is decreasing, or vice versa if the ramp is increasing. To try avoid effects such as these, the lightcurve analysis is often performed separately for a number of different ramp functions, and the one that minimises the model residuals or maximises an approximation to the Bayesian model evidence, such as the negative Bayesian information criterion (Schwarz 1978), is retained (e.g. Stevenson et al. 2012).

5   GAUSSIAN PROCESSES FOR IRAC SYSTEMATICS

In this paper, we present an alternative method for treating intrapixel sensitivity variations and the ramp effect in IRAC lightcurves based on Gaussian processes (GPs). The first application of GP models to transit lightcurves was made by Gibson et al. (2012b), with subsequent applications in Gibson et al. (2012a, 2013a,b), Evans et al. (2013), and Gibson (2014), to which the reader is referred for further details.
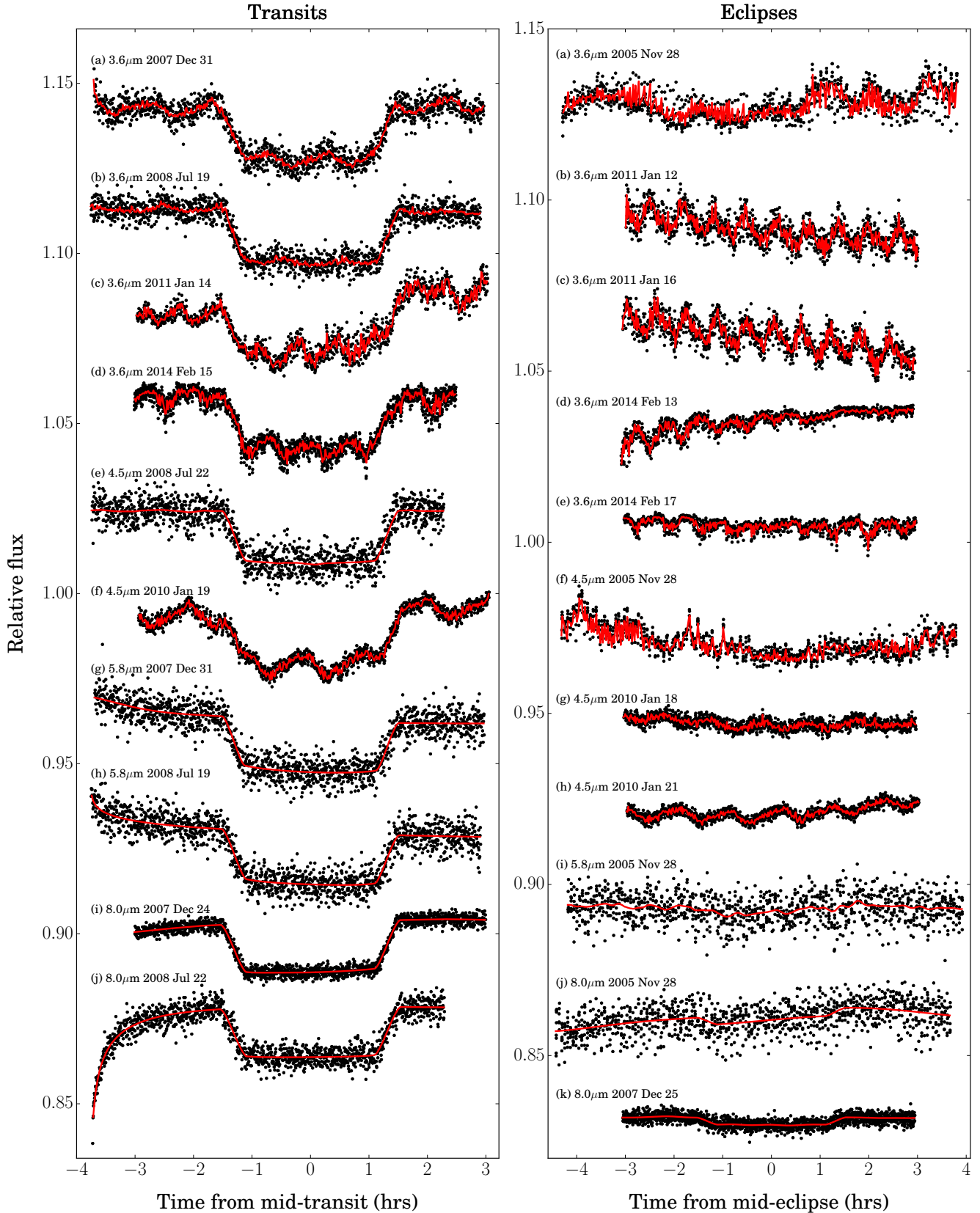
Figure 1. Raw lightcurves obtained for the 10 transits (left column) and 11 eclipses (right column) analysed in this study. Red lines show best-fit GP models, which are described in Section 5. The lightcurves shown in this plot have been binned in time, which was done to make the GP model fitting computationally tractable (Section 5.3). The plotted GP models have been further binned into 1 minute bins and vertical offsets have been applied to each lightcurve for clarity.

Formally, a GP is defined as a collection of data points, any subset of which has a multivariate normal distribution (e.g. Rasmussen & Williams 2006). Indeed, this assumption is regularly made for transit lightcurve analyses published in the literature, even if it is not stated explicitly. It is equivalent to assuming that what we measure is some underlying signal, which is the combination of the astrophysical signal of interest and additional systematic terms, plus uncorrelated Gaussian noise. If we are in possession of a model μ that describes the underlying signal, the probability of measuring a specific dataset $\mathbf{d} = \{d_1, d_2, \ldots, d_N\}$ with Gaussian errorbars $\boldsymbol{\sigma_w} = \{\sigma_{w,1}, \sigma_{w,2}, \ldots, \sigma_{w,N}\}$ for a set of model parameters $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_M\}$ is given by:

$$p(\mathbf{d}|\boldsymbol{\alpha}) = \prod_{i=1}^{N} \frac{1}{\sigma_{w,i}\sqrt{2\pi}} \exp\left[-\frac{(d_i - \mu_i)^2}{2\sigma_{w,i}^2}\right] = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ , \quad (1)$$

where $\mathcal{N}$ denotes a multivariate normal distribution and $\boldsymbol{\Sigma} = \text{diag}[\boldsymbol{\sigma_w}]$ is the diagonal covariance matrix. For fixed errorbars, optimising the log likelihood is therefore identical to the familiar practice of minimising the χ^2 statistic, since:

$$\ln \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{1}{2}\chi^2 - \frac{1}{2}\sum_{i=1}^{N}\ln\sigma_{w,i} - \frac{1}{2}N\ln 2\pi \ . \quad (2)$$

Note, however, that Equation 2 requires us to build all of our information about the underlying signal into the deterministic mean function $\boldsymbol{\mu}$. This is not desirable if, as is commonly the case, the systematics are poorly understood from first principles and an explicit functional form is not available to describe them.

An alternative option is to incorporate the systematics into our model by allowing for nonzero off-diagonal entries in the covariance matrix of the likelihood function, such that:

$$\ln p(\mathbf{d}|\boldsymbol{\alpha}, \boldsymbol{\gamma}) = \ln \mathcal{N}(\boldsymbol{\mu}, \mathbf{K} + \boldsymbol{\Sigma}) \ , \quad (3)$$

where K_{ij} gives the covariance between the ith and jth data points, and $\boldsymbol{\gamma}$ are the parameters that control the behaviour of the covariance. Although we no longer have to provide an explicit functional form for the systematics contribution, we must now specifiy a kernel function to populate the entries of the covariance matrix $\mathbf{K}$. However, by modelling the statistical covariance between data points rather than the deterministic systematics signal directly, it is possible to capture a broad range of behaviours with relatively few free parameters. Thus, GP models are simultaneously parsimonious and flexible.

Before proceeding to describe the specific mean functions and covariance kernels adopted in the current study in Sections 5.1 and 5.2, it is worth also pointing out the fact that GPs have the desirable property of automatically implementing the principle of Occam's razor. This can be seen if we expand Equation 3 into its constituent terms:

$$\ln p(\mathbf{d}|\boldsymbol{\alpha}, \boldsymbol{\gamma}) = -\frac{1}{2}\mathbf{r}^{\mathrm{T}}(\mathbf{K}+\boldsymbol{\Sigma})^{-1}\mathbf{r} - \frac{1}{2}\ln|\mathbf{K}+\boldsymbol{\Sigma}| - \frac{1}{2}N\ln 2\pi \ , \quad (4)$$

where $\mathbf{r}$ is a vector containing the model residuals, with ith term given by $r_i = d_i - \mu_i$. The first term on the righthand side, $-\frac{1}{2}\mathbf{r}^{\mathrm{T}}(\mathbf{K}+\boldsymbol{\Sigma})^{-1}\mathbf{r}$, serves as a goodness-of-fit term. For a given covariance matrix, it increases as the residuals become smaller, rewarding mean functions $\boldsymbol{\mu}$ that match the data well. The second term, $-\frac{1}{2}\ln|\mathbf{K}+\boldsymbol{\Sigma}|$, can be thought of as a complexity penalty. This is because increasing the com-

Table 3. Nonlinear limb darkening coefficients.

	Channels (μm)			
	3.6	4.5	5.8	8.0
c_1	0.5564	0.4614	0.4531	0.4354
c_2	−0.5462	−0.4277	−0.5119	−0.6067
c_3	0.4315	0.3362	0.4335	0.5421
c_4	−0.1368	−0.1074	−0.1431	−0.1816

plexity of a model is equivalent to assigning similar probabilities to an increasing diversity of functions. In other words, as the model complexity increases, the likelihood function becomes less sharply peaked near the mean $\boldsymbol{\mu}$ and the probability mass of the model becomes more diffusely spread throughout the function space. This is precisely what happens as the term $-\frac{1}{2}\ln|\mathbf{K}+\boldsymbol{\Sigma}|$ decreases, in effect penalising model complexity. Finally, the third term, $-\frac{1}{2}N\ln 2\pi$, remains constant for a given dataset. The balance between the first two terms of Equation 4 therefore ensures that the probability mass of the model is distributed over the parameter space in a manner that optimises the trade-off between goodness-of-fit and model complexity.

5.1 Mean functions

The mean function $\boldsymbol{\mu}$ defines the model for the astrophysical signal with well-understood form; namely, a primary transit or secondary eclipse. For this purpose, we adopt the analytic transit functions of Mandel & Agol (2002). We set the orbital eccentricity to zero, consistent with observational evidence (e.g. Pont et al. 2011). We fix the orbital period to $P = 3.52474859$ day (Knutson et al. 2007b).

For the primary transits, the mean function parameters that are allowed to vary are the radius ratio $R_p/R_\star$, normalised semimajor axis $a/R_\star$, impact parameter $b = a\cos i/R_\star$, and transit mid-time T_{mid}, such that $\boldsymbol{\alpha} = \{R_p/R_\star, a/R_\star, b, T_{\mathrm{mid}}\}$ in Equation 4. Stellar limb darkening is treated using the nonlinear law of Claret (2004), with coefficients fixed to those provided by Hayek et al. (2012) which were obtained specifically for the IRAC bandpasses using a 3D stellar model for HD 209458 (Table 3).

For the secondary eclipses, only the eclipse depth $F_p/F_\star$ and eclipse mid-time T_{mid} are allowed to vary, such that $\boldsymbol{\alpha} = \{F_p/F_\star, T_{\mathrm{mid}}\}$. The remaining mean function parameters are fixed to the values published by Torres et al. (2008), namely, $R_p/R_\star = 0.121$, $a/R_\star = 8.76$ and $b = 0.507$.

5.2 Covariance kernels

Entries of the covariance matrix $\mathbf{K}$ are constructed using a kernel function, such that $K_{ij} = k(\mathbf{v_i}, \mathbf{v_j})$, where $\mathbf{v_i}$ and $\mathbf{v_j}$ are vectors of inputs associated with the ith and jth data points, respectively. By inputs, we refer to variables that correlate with the measured signal – these are typically the same variables that would be used for standard polynomial systematics decorrelations. Further discussion of common

GP kernels, such as the squared exponential and Matérn kernels, can be found in Gibson et al. (2012b).

In this study, we parameterise the entries of the covariance matrix $\mathbf{K}$ for the 3.6μm and 4.5μm channels as the sum of two kernels: a squared exponential kernel with the centroid xy coordinates as inputs and a Matérn $\nu = 3/2$ kernel with time t as the input. Writing this out explicitly, the combined kernel is given by:

$$k\left(\boldsymbol{v_i}, \boldsymbol{v_j}\right) = k_{xy} + k_t, \qquad (5)$$

where:

$$k_{xy} = A_{xy}^2 \exp\left[-\left(\frac{x_i - x_j}{L_x}\right)^2 - \left(\frac{y_i - y_j}{L_y}\right)^2\right], \qquad (6)$$

$$k_t = A_t^2 \left[1 + \frac{t_i - t_j}{L_t}\sqrt{3}\right] \exp\left[-\left(\frac{t_i - t_j}{L_t}\right)\sqrt{3}\right], \quad (7)$$

such that $\boldsymbol{\gamma} = \{A_{xy}, L_x, L_y, A_t, L_t\}$ in the notation of Equation 4. The squared exponential component of this kernel accounts for the smooth spatial variations in pixel sensitivities that dominate the systematics of the $3.6\,\mu$m and $4.5\,\mu$m channels, while the Matérn component accounts for any residual correlated noise in the lightcurve.

For the 5.8μm and 8.0μm channel lightcurves, we used a squared exponential kernel to model the dominant ramp effect, with form given by:

$$k_\tau = A_\tau^2 \exp\left[-\left(\frac{\tau_i - \tau_j}{L_\tau}\right)^2\right], \qquad (8)$$

where $\tau = \ln(t + h)$ is logarithmic time t, and h is a parameter that can be inferred from the data (see below). Parameterising the covariance with k_τ allows us to capture the dominant behaviour of the ramp effect; namely, a steep initial gradient followed by a levelling off of the measured flux (Figure 1). However, we stress that by parameterising the covariance between data points according to Equation 8, we are not constraining the systematics to be monotonically increasing or decreasing in time. This is a valuable property of GPs given that the ramp effect is not necessarily strictly monotonic, with an overshoot effect identified in a number of datasets (e.g. Knutson et al. 2012). As with the other channels, we also include a time-dependent Matérn $\nu = 3/2$ kernel k_t to account for residual correlations in the lightcurve that may not be related to the ramp. Therefore, the final kernel is given by:

$$k\left(\boldsymbol{v_i}, \boldsymbol{v_j}\right) = k_\tau + k_t, \qquad (9)$$

with covariance parameters $\boldsymbol{\gamma} = \{A_\tau, L_\tau, h, A_t, L_t\}$.

The covariance kernels outlined above allow for systematics treatments that are at least as versatile as others used in the literature. For instance, the k_{xy} component of Equation 5 is similar in concept to the 2D Gaussian correction developed by Ballard et al. (Section 4.1). Similarly, the k_τ component of Equation 9 is reminiscent of the logarithmic time polynomials used in other published studies (Section 4.2). However, as it is only the covariance between data points that is parameterised in terms of logarithmic time, the underlying signal itself need not be monotonically increasing or decreasing. Furthermore, by parameterising the covariance rather than systematics signal directly, the GP model is capable of marginalising over a broader range of function space with relatively few tunable parameters.

Before proceeding, we highlight the fact that the covariance kernels given by Equations 5–9 assume that the input variables are noise-free. While this is a reasonable assumption for time t, it is not necessarily the case for the centroid coordinates x and y. Indeed, the undersampled nature of the IRAC PSF makes the centroid estimates particularly susceptible to shot noise of individual pixels. However, due to the fact that this noise is white, and because in practise a small patch of a single pixel is densely sampled by the PSF over the course of a few hours, we expect its effect to be averaged out. For this reason, and in line with other analyses of IRAC lightcurves, we do not explicitly account for the noise of x and y. In future work, however, it may be worth considering a more explicit treatment of noisy inputs (e.g. Goldberg et al. 1998; Mchutchon & Rasmussen 2011), or simply smoothing noisy inputs before feeding them to the GP (e.g. Gibson et al. 2012b).

5.3 Lightcurve binning

GP models become computationally intractable for datasets with $N \gtrsim 10^3$ data points. This is due to the need to factorise the $N \times N$ covariance matrix when evaluating the $-\frac{1}{2}\mathbf{r}^{\mathrm{T}}(\mathbf{K} + \boldsymbol{\Sigma})^{-1}\mathbf{r}$ term and computing the determinant $-\frac{1}{2}\ln|\mathbf{K} + \boldsymbol{\Sigma}|$ for each log likelihood evaluation (Equation 4). Our code implements this using Cholesky factorisation, which has a computational cost scaling as $\mathcal{O}\left(N^3\right)$. To apply GPs to IRAC datasets, most of which consist of $N > 10^4$ data points, we bin the fluxes and centroid xy coordinates in time prior to fitting. Binning factors were chosen according to the format of individual lightcurves, such that the time interval between successive binned points was $\lesssim 15\,\mathrm{sec}$ and the number of binned points per lightcurve was $N_b = 1000\text{–}1500$. Lightcurves obtained in full array mode for Program 461 were binned by a factor of two, giving a median cadence of about $17\,\mathrm{sec}$; lightcurves obtained in subarray mode for Programs 40280 and 60021 with frame times of $0.4\,\mathrm{sec}$ were binned by a factor of 32, giving a median cadence of about $14\,\mathrm{sec}$; lightcurves obtained in subarray mode for Program 60021 with frame times of $0.1\,\mathrm{sec}$ were binned by a factor of 128, giving a median cadence of about $17\,\mathrm{sec}$; lightcurves obtained in subarray mode for Program 20523 with frame times of $0.1\,\mathrm{sec}$ were only binned by a factor of 32, due to the sparser sampling of the lightcurve as the telescope was constantly repointed during the observations, resulting in a median cadence of about $5\,\mathrm{sec}$. Details are given in Table 2.

The obvious drawback of binning the lightcurves in time is that we lose information on the timescales of our bin sizes. For instance, the xy centroid coordinates can vary coherently over timescales $\lesssim 15\,\mathrm{sec}$ (e.g. Stevenson et al. 2012). However, this should not affect our results significantly, as none of the astrophysical quantities of interest vary over the bin timescales, nor does the information content of the transit lightcurve degrade significantly as we reduce the time resolution to $\lesssim 15\,\mathrm{sec}$. Furthermore, high-frequency systematics should mostly average out given the large number of binned data points for each tunable model parameter. Any correlations that remain will be accounted for by the time-dependent Matérn $\nu = 3/2$ kernel k_t in our model (Equations 5 and 9).

6   LIGHTCURVE FITTING

Lightcurves were fit individually by marginalising over the model posterior distribution using Markov chain Monte Carlo (MCMC), in order to quantify the degeneracies between the astrophysical parameters of interest and instrumental systematics. Following Bayes theorem, the model posterior distribution is given by $p(\boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{d}) \propto p(\mathbf{d}|\boldsymbol{\alpha}, \boldsymbol{\gamma})\, p(\boldsymbol{\alpha})\, p(\boldsymbol{\gamma})$, where $p(\mathbf{d}|\boldsymbol{\alpha}, \boldsymbol{\gamma})$ is the GP likelihood given by Equation 4, and $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\gamma})$ are the priors on the mean function and covariance parameters, respectively. We adopted uniform priors for the mean function parameters $\boldsymbol{\alpha}$ and covariance length scales L_i. For the covariance amplitudes A_i, Gamma distribution priors of the form $p(A_i) = \mathrm{Gam}(1, 100) \propto \exp[-100 A_i]$ for $i = \{t, xy, \tau\}$ were adopted. The latter give decreasing probability to increasing covariance amplitudes, encouraging the GP to reduce the covariance amplitude unless justified by the data. The white noise level σ_w was included as a free parameter in each model, with a uniform prior. The ability to inflate the statistical errorbars above the formal shot noise floor provides the models with some additional flexibility for dealing with high-frequency noise that may be present in the data, without having to reduce the correlation length scales L_i of the covariance kernels to unreasonably small values.

With the posterior distributions defined, the model fitting for each lightcurve proceeds as follows. Values for the model parameters were drawn randomly from the model prior, i.e. $p(\boldsymbol{\alpha})\, p(\boldsymbol{\gamma})$. With this as a starting point, Equation 4 was optimised using the Nelder-Mead simplex algorithm (Nelder & Mead 1965) to obtain maximum likelihood estimates (MLEs) for the parameters. A short Metropolis-Hastings MCMC chain (Metropolis et al. 1953; Hastings 1970) of 1000 steps was initiated at the MLE, with step sizes pretuned to give acceptance rates of 20–40 %. The median chain values were then used as the starting location for a second MLE optimisation. In practise, the randomness introduced by the short MCMC chains helped prevent the MLE optimisations from getting trapped at local maxima of the likelihood surface, thus increasing the chance of locating the global likelihood maximum. To further increase this probability, the entire process was repeated ten times, each time from a different random starting point. This was done separately for each lightcurve produced using the different photometric aperture sizes (Section 3). The photometric reduction giving the lowest scatter in the residuals was then selected for the remaining analysis.

Before commencing the final MCMC chains, the covariance parameters were fixed to their MLE values, which are reported in Table A1. This approach – which is often referred to as "type-II maximum likelihood" (for further discussion see Gibson et al. 2012b) – allows the expensive $\mathcal{O}\left(N^3\right)$ covariance matrix factorisation required for the GP likelihood evaluation (Equation 4) to be performed only once at the beginning of the chain. Subsequent steps only cost $\mathcal{O}\left(N^2\right)$, resulting in much faster computations. The disadvantage is that by fixing the covariance parameters $\boldsymbol{\gamma}$, they are not marginalised over. In effect, this imposes an artificial restriction on the range of systematics functions that are explored by the GP model. Consequently, there may be degeneracies between the planet signal and systematics that are not fully incorporated into the final uncertainties

for the planet parameters $\boldsymbol{\alpha}$ presented here. For example, in their re-analysis of the NICMOS transmission spectrum for HD 189733b, Gibson et al. (2012b) found uncertainties that were up to ~ 1.5 larger when the covariance parameters were allowed to vary in the marginalisation compared to when they were fixed to their MLE values. Therefore, the uncertainties we report in this study should be considered lower limits to the true uncertainties.

It should be emphasised, however, that fixing the covariance parameters is quite different to fixing the parameters of an explicit functional model for the systematics. Instead, fixing the covariance parameters is somewhat analogous to selecting a family of parametric models for the systematics, as they control the high-level properties of the function space spanned by the GP model. Rather than selecting from a handful of distinct parametric models, the GP model offers access to a continuum of possible functions. By using covariance parameters that optimise the GP likelihood, this continuum is narrowed in a principled manner. To compare with the bilinear interpolation method for pixel mapping used by Stevenson et al. (2012) for instance, optimising the covariance length scales L_x and L_y is similar to choosing the optimal grid spacing for the interpolation knots. Treating our dataset as a GP, we have the advantage of being able to do this in the context of a self-consistent probabilistic model, by maximising the likelihood function with respect to the unknown parameters using a numerical optimiser.

Having fixed the covariance parameters, an initial chain of 10^5 steps was run with the planet parameters allowed to vary and step sizes again pretuned to ensure acceptance rates of 20–40 %. The first 5×10^4 steps were discarded as burn-in. An additional four chains were then run for 10^5 steps each, with starting parameter values drawn randomly from normal distributions centred on the mean values of the first chain. The width of the normal distributions were taken to be five times the standard deviation of the first chain, to ensure the starting locations were well-dispersed in parameter space. After discarding the first 5×10^4 burn-in steps of these chains, the Gelman-Rubin statistics for each parameter were calculated (Gelman & Rubin 1992). For all lightcurves, these were found to be well within 1% of unity, consistent with the chains having reached stable states. Finally, the five independent chains were combined into a single chain, giving 2.5×10^5 samples from the posterior distribution.

7   RESULTS

Results of the primary transit and secondary eclipse MCMC analyses are given in Table 4. Best-fit models are overplotted on the raw lightcurves in Figure 1 and the corrected lightcurves with residuals are shown in Figure 2.

7.1   Transmission ($R_p/R_\star$)

We find good agreement in the inferred parameters across epochs for the majority of the transit lightcurves. The only exceptions to this are the 2011 Jan 14 and 2014 Feb 15 transits measured in the 3.6μm channel. For the 2011 lightcurve we obtain a value for $R_p/R_\star$ that is $> 6\sigma$ discrepant relative to those obtained for the 2007 Dec 31 and 2008 Jul 19 3.6μm
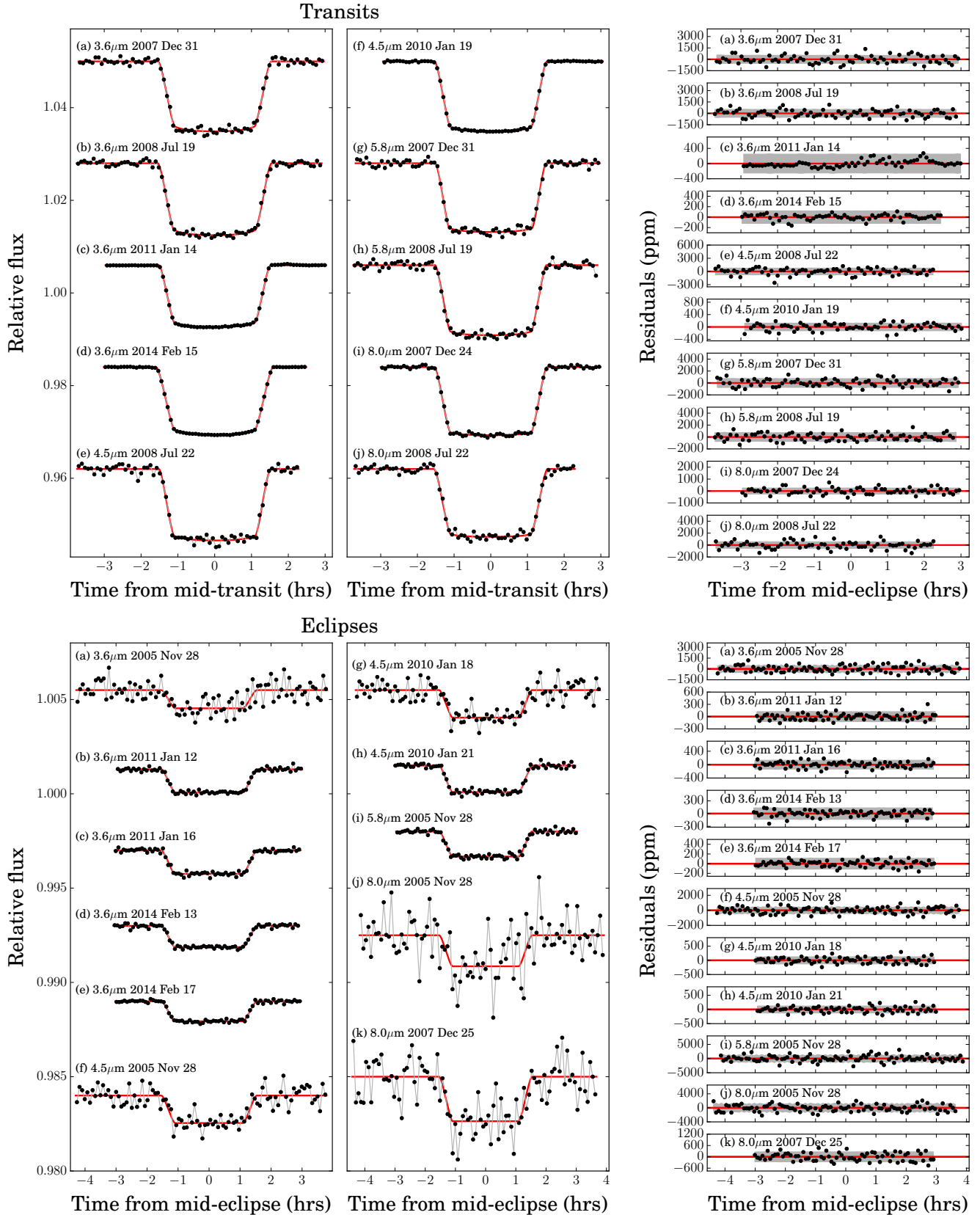
Transits

Eclipses

Figure 2. Corrected primary transit (top) and secondary eclipse (bottom) lightcurves obtained by dividing the raw lightcurves by the systematics components of the best-fit GP models. Model residuals are shown in the rightmost column, with grey shaded regions indicating the inferred white noise values σ_w.

Table 4. Results of MCMC primary transit lightcurve analyses. Quoted values are the chain medians, with uncertainties corresponding to the ranges either side of the medians that contain 34% of the chain samples. Orbital inclination values i are derived from the corresponding impact parameter $b = a \cos i/R_\star$ and normalised semimajor axis $a/R_\star$ values. Brightness temperatures T_b are derived from the measured eclipse depths $F_p/F_\star$ assuming an ATLAS stellar model (Kurucz 1979, 1993) for HD 209458 and integrating over the IRAC bandpasses.

			Transits			
Channel (μm)	Date	T_{mid} (BJD$_{\mathrm{UTC}}$ − 2450000)	$R_p/R_\star$	$a/R_\star$	b	i (deg)
3.6	2007 Dec 31	$4465.63711^{+0.00027}_{-0.00027}$	$0.12077^{+0.00085}_{-0.00084}$	$8.72^{+0.26}_{-0.23}$	$0.527^{+0.035}_{-0.044}$	$86.54^{+0.39}_{-0.33}$
	2008 Jul 19	$4666.54742^{+0.00021}_{-0.00021}$	$0.12220^{+0.00062}_{-0.00062}$	$8.89^{+0.33}_{-0.26}$	$0.482^{+0.044}_{-0.062}$	$86.89^{+0.50}_{-0.39}$
	2011 Jan 14	$5575.93102^{+0.00030}_{-0.00030}$	$0.11354^{+0.00085}_{-0.00087}$	$8.77^{+0.33}_{-0.33}$	$0.525^{+0.048}_{-0.056}$	$86.57^{+0.48}_{-0.46}$
	2014 Feb 15	$6703.85250^{+0.00011}_{-0.00011}$	$0.11919^{+0.00032}_{-0.00032}$	$8.13^{+0.10}_{-0.10}$	$0.590^{+0.013}_{-0.014}$	$85.84^{+0.14}_{-0.14}$
4.5	2008 Jul 22	$4670.07290^{+0.00029}_{-0.00029}$	$0.12199^{+0.00094}_{-0.00091}$	$9.31^{+0.35}_{-0.30}$	$0.423^{+0.057}_{-0.079}$	$87.39^{+0.56}_{-0.45}$
	2010 Jan 19	$5216.40564^{+0.00007}_{-0.00007}$	$0.12099^{+0.00029}_{-0.00029}$	$8.89^{+0.06}_{-0.07}$	$0.493^{+0.011}_{-0.010}$	$86.82^{+0.08}_{-0.10}$
5.8	2007 Dec 31	$4465.63663^{+0.00032}_{-0.00031}$	$0.12007^{+0.00248}_{-0.00265}$	$9.20^{+0.36}_{-0.35}$	$0.435^{+0.066}_{-0.082}$	$87.29^{+0.59}_{-0.53}$
	2008 Jul 19	$4666.54744^{+0.00033}_{-0.00033}$	$0.11880^{+0.00284}_{-0.00272}$	$9.27^{+0.38}_{-0.37}$	$0.427^{+0.071}_{-0.091}$	$87.36^{+0.64}_{-0.56}$
8.0	2007 Dec 24	$4458.58730^{+0.00013}_{-0.00013}$	$0.12007^{+0.00114}_{-0.00114}$	$8.78^{+0.11}_{-0.11}$	$0.513^{+0.017}_{-0.019}$	$86.65^{+0.16}_{-0.15}$
	2008 Jul 22	$4670.07243^{+0.00022}_{-0.00022}$	$0.11991^{+0.00073}_{-0.00073}$	$8.67^{+0.20}_{-0.18}$	$0.527^{+0.027}_{-0.032}$	$86.52^{+0.28}_{-0.26}$

		Eclipses		
Channel (μm)	Date	T_{mid} (BJD$_{\mathrm{UTC}}$ − 2450000)	$F_p/F_\star$ (%)	T_b (K)
3.6	2005 Nov 28	$3702.52741^{+0.00476}_{-0.00457}$	$0.093^{+0.033}_{-0.034}$	1447^{+163}_{-165}
	2011 Jan 12	$5574.16955^{+0.00173}_{-0.00190}$	$0.122^{+0.011}_{-0.011}$	1591^{+49}_{-49}
	2011 Jan 16	$5577.69695^{+0.00204}_{-0.00193}$	$0.124^{+0.014}_{-0.014}$	1599^{+65}_{-65}
	2014 Feb 13	$6702.09332^{+0.00397}_{-0.00400}$	$0.112^{+0.016}_{-0.017}$	1545^{+76}_{-80}
	2014 Feb 17	$6705.61556^{+0.00136}_{-0.00128}$	$0.106^{+0.007}_{-0.008}$	1515^{+34}_{-37}
4.5	2005 Nov 28	$3702.52942^{+0.00169}_{-0.00191}$	$0.145^{+0.018}_{-0.017}$	1474^{+72}_{-69}
	2010 Jan 18	$5214.64693^{+0.00085}_{-0.00079}$	$0.140^{+0.014}_{-0.014}$	1455^{+57}_{-56}
	2010 Jan 21	$5218.16915^{+0.00147}_{-0.00138}$	$0.133^{+0.011}_{-0.011}$	1426^{+47}_{-48}
5.8	2005 Nov 28	$3702.53033^{+0.00457}_{-0.00664}$	$0.142^{+0.059}_{-0.058}$	1297^{+225}_{-220}
8.0	2005 Nov 28	$3702.53461^{+0.00322}_{-0.00344}$	$0.225^{+0.064}_{-0.063}$	1433^{+219}_{-216}
	2007 Dec 25	$4460.35170^{+0.00100}_{-0.00107}$	$0.215^{+0.012}_{-0.012}$	1397^{+41}_{-43}

lightcurves. This lightcurve has been classified as a failed observation by the *Spitzer* Science Center due to the presence of high-frequency noise of unknown origin during the second half of the transit. For the 2014 lightcurve, we obtain values for $a/R_\star$ and b that are both $\sim 2\sigma$ discrepant relative to the values inferred for the 2007 and 2008 lightcurves, while the value inferred for $R_p/R_\star$ is $> 4\sigma$ discrepant. The source of this disagreement is not clear. We also performed the lightcurve fitting using polynomial xy decorrelations and the PLD method of Deming et al. (2014), but these analyses gave similarly discrepant results. We therefore suspect that there is either an issue with the data itself or our photometric reduction for this particular lightcurve. For these reasons, we do not consider the 2011 and 2014 3.6μm transit lightcurves any further in this paper. However, we consider our analyses for the eclipses in these lightcurves to be more robust, as they gave results that are consistent with those obtained at other epochs.

HD209458b transmission

HD209458b emission
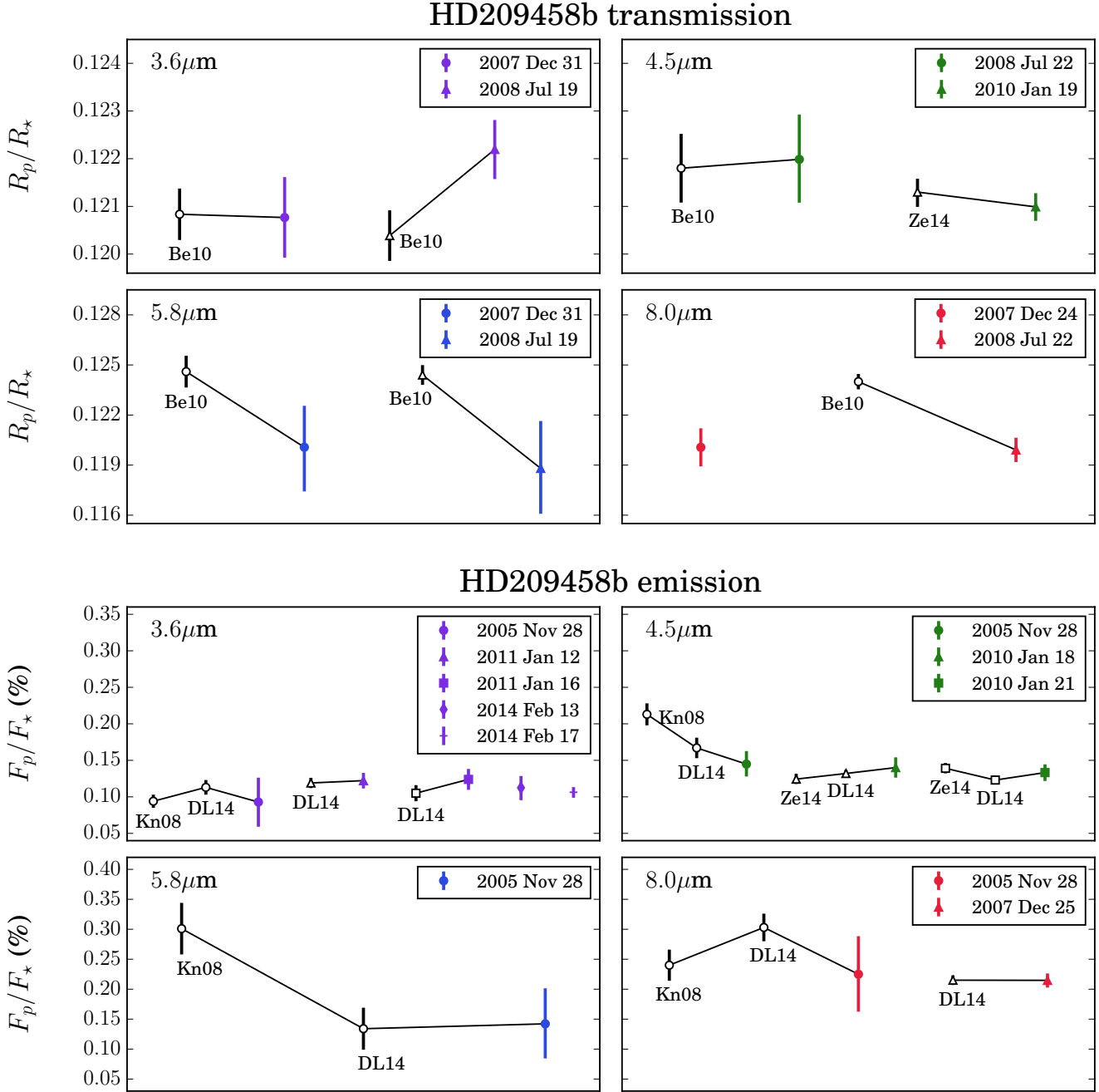
Figure 3. Comparison of transmission $R_p/R_\star$ (top) and emission $F_p/F_\star$ (bottom) results obtained in the present study with those published in the literature. Wavelength channels are indicated in the top left corner of each axis. Independent analyses of the same dataset are linked by solid black lines: filled symbols show the values obtained in the current study and unfilled symbols show values obtained by other authors. The latter are labelled using the same abbreviations adopted in Table 1.

For the remaining eight transit lightcurves, inferred values for $R_p/R_\star$ are shown in Figure 3, along with values previously published in the literature. In the $3.6\,\mu$m channel, we find $R_p/R_\star = 0.12077^{+0.00085}_{-0.00084}$ for the 2007 Dec 31 lightcurve, which is in agreement with the value of $R_p/R_\star = 0.120835 \pm 0.00054$ reported by Beaulieu et al. (2010). We obtain a somewhat higher value of $R_p/R_\star = 0.12220^{+0.00062}_{-0.00062}$ for the 2008 Jul 19 lightcurve, compared with the value of $R_p/R_\star = 0.120387 \pm 0.00053$ obtained by Beaulieu et al.

Our results for the $4.5\,\mu$m lightcurves are in good agree-

ment both with each other and with the values previously published by Beaulieu et al. (2010) and Zellem et al. (2014).

For the $5.8\,\mu$m 2007 Dec 31 and 2008 Jul 19 lightcurves, we find $R_p/R_\star = 0.12007^{+0.00248}_{-0.00265}$ and $R_p/R_\star = 0.11880^{+0.00284}_{-0.00272}$, respectively. Although these values are consistent with each other, they are 1.7σ and 1.9σ lower, respectively, than the corresponding values of $R_p/R_\star = 0.1246 \pm 0.00095$ and $R_p/R_\star = 0.1244 \pm 0.00059$ obtained by Beaulieu et al. (2010). In particular, our uncertainties are ~ 2.5–3 times larger than those of Beaulieu et al. This is most likely

due to the flexibility of the GP models allowing broader ranges of function space to be marginalised over, which in turn maps out broader degeneracies between the transit parameters and the systematics contributions.

For the 8.0μm channel, we find good agreement between our inferred parameters for the 2007 Dec 24 and 2008 Jul 22 lightcurves, with $R_p/R_\star = 0.12007^{+0.00114}_{-0.00114}$ and $R_p/R_\star = 0.11991^{+0.00073}_{-0.00073}$, respectively. For the 2008 lightcurve, our value is 4.7σ lower than the value of $R_p/R_\star = 0.1240 \pm 0.00046$ published by Beaulieu et al. (2010).

7.2 Emission ($F_p/F_\star$)

As with the transmission measurements, we find consistent results across epochs and wavelength channels for the inferred emission values (Figure 3). Our results are mostly in agreement with values previously published in the literature, but with typically larger uncertainties. There are two notable exceptions. Firstly, for the 4.5μm 2005 Nov 28 lightcurve, we obtain $F_p/F_\star = 0.145^{+0.018}_{-0.017}\%$, which is 2.9$\sigma$ lower than the value of $F_p/F_\star = 0.213 \pm 0.015\%$ published by Knutson et al. (2008), but in agreement with the value of $F_p/F_\star = 0.134 \pm 0.035\%$ published by Diamond-Lowe et al. (2014) for the same lightcurve. Our revision brings the value into good agreement with those obtained in the same wavelength channel at different epochs, both in the current study and by Zellem et al. (2014). Secondly, for the 2005 Nov 28 5.8μm lightcurve, we obtain $F_p/F_\star = 0.142^{+0.059}_{-0.058}\%$, which is 2.3$\sigma$ lower than the value of $F_p/F_\star = 0.310 \pm 0.043\%$ published by Knutson et al. (2008), but in agreement with the value of $F_p/F_\star = 0.134 \pm 0.035\%$ published by Diamond-Lowe et al. (2014). Finally, we note that unlike Diamond-Lowe et al. (2014), who obtained $F_p/F_\star$ values for the 2005 Nov 28 and 2007 Dec 25 8.0μm lightcurves that conflicted at the 3.6σ level, our values are in good agreement with each other due to the relatively large uncertainty we obtain for the 2005 Nov 28 eclipse depth. We therefore favour the more conservative uncertainty estimate provided by our GP analysis in this case.

7.3 Orbital parameters ($a/R_\star$, b)

The normalised semimajor axis $a/R_\star$ and orbital inclination i values recovered from the primary lightcurve analyses are plotted in Figure 4. Note the clear correlation between both parameters, i.e. higher values of $a/R_\star$ are associated with higher values of i, and vice versa. This reflects the fact that these parameters exert opposing influences on the transit duration.

We expect $a/R_\star$ and i remain constant in time across the different wavelength channels. Computing the weighted arithmetic mean across epochs, we obtain $a/R_\star = 8.87 \pm 0.05$, $b = 0.499 \pm 0.008$ and $i = 86.78 \pm 0.07$ deg. We caution that the quoted uncertainties for the weighted means are likely to be underestimated, as they are calculated by combining multiple measurements under the assumption that the errorbars are normally distributed, which is not necessarily true. Nonetheless, these values are consistent at the $\sim 1\sigma$ level with the values of $a/R_\star = 8.77 \pm 0.07$ and $i = 86.76 \pm 0.10$ deg obtained by Beaulieu et al. (2010), and $a/R_\star = 8.810^{+0.064}_{-0.069}$ and $i = 86.69^{+0.09}_{-0.10}$ deg obtained by Zellem et al. (2014).
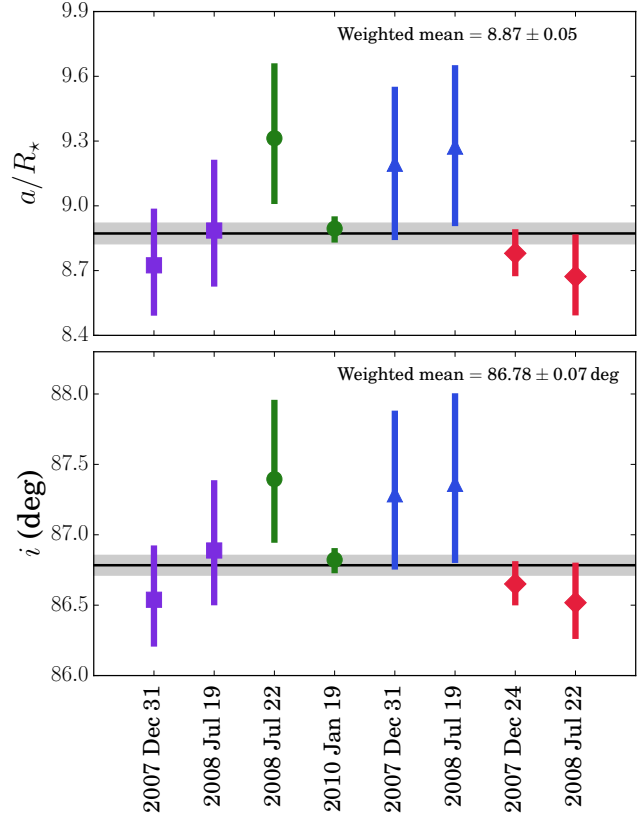
Figure 4. Normalised semimajor axis $a/R_\star$ and orbital inclination values obtained from the MCMC analyses. Lightcurve epochs are labelled along the horizontal axis, with different marker symbols for each wavelength channel: purple squares for 3.6μm, green circles for 4.5μm, blue triangles for 5.8μm and red diamonds for 8.0μm. Weighted mean values are labelled on each axis, and shown as horizontal black lines with grey shaded regions indicating the corresponding 1σ uncertainties.

Indeed, the values obtained for $a/R_\star$ and $b = a/R_\star \cos i$ in the present study are overall more consistent across wavelengths and epochs than previously published values. For instance, the $a/R_\star$ and i values reported by Beaulieu et al. (2010) for the 8.0μm channel differ from most of the values they obtain in the other channels by 3–5σ. The relative consistency of the orbit parameters derived in the current study therefore offers further evidence that the GP modelling approach is doing a good job of accounting for the lightcurve systematics and providing realistic parameter uncertainties.

7.4 Ephemeris

Using the transit and eclipse mid-times listed in Table 4, we computed the ephemeris assuming a constant orbital period P, with the linear relation $T_0 = T_{\mathrm{mid}}(n) + nP$, where n is the number of orbital periods since the reference epoch T_0. Eclipse mid-times were treated as occurring precisely $0.5P$ after the immediately-preceding transit. Before performing the fit, we converted the BJD timestamps from Coordinated Universal Time (UTC) to the Barycentric Dynamical Time (TDB) standard, as recommended by Eastman et al. (2010). To do this, we added the appropriate number of leap seconds to the UTC timestamps, namely: 64.184 sec for the

Table 5. Results of the ephemeris fit to transit and eclipse mid-times assuming a constant orbital period and zero eccentricity. As discussed in Section 7.4, fits were performed both with and without the 2010 Jan 19 transit mid-time. As discussed in the text, we favour the ephemeris obtained with the 2010 Jan 19 transit excluded from the fit.

			$O - C$ residual (min)	
Date	Signal	Channel (μm)	Including 2010 transit	Excluding 2010 transit
2005 Nov 28	Eclipse	3.6	-6.6 ± 6.7	-1.8 ± 6.7
		4.5	-3.7 ± 2.6	1.1 ± 2.7
		5.8	-2.4 ± 8.1	2.4 ± 8.1
		8.0	3.8 ± 4.8	8.6 ± 4.8
2007 Dec 24	Eclipse	8.0	-0.5 ± 0.2	-0.1 ± 0.2
2007 Dec 25	Transit	8.0	2.4 ± 1.5	2.8 ± 1.5
2007 Dec 31	Transit	3.6	-0.1 ± 0.4	0.3 ± 0.4
		5.8	-0.7 ± 0.5	-0.4 ± 0.5
2008 Jul 19	Transit	3.6	0.5 ± 0.3	-0.3 ± 0.3
		5.8	0.6 ± 0.5	-0.2 ± 0.5
2008 Jul 22	Transit	4.5	1.6 ± 0.4	0.8 ± 0.4
		8.0	0.9 ± 0.3	0.1 ± 0.3
2010 Jan 18	Eclipse	4.5	5.1 ± 1.2	1.1 ± 1.3
2010 Jan 19	Transit	4.5	-0.2 ± 0.2	-4.7 ± 0.5
2010 Jan 21	Eclipse	4.5	1.5 ± 2.1	-2.5 ± 2.1
2011 Jan 12	Eclipse	3.6	4.6 ± 2.6	-1.5 ± 2.7
2011 Jan 16	Eclipse	3.6	8.4 ± 2.9	2.3 ± 3.0
2014 Feb 13	Eclipse	3.6	16.2 ± 5.8	3.7 ± 5.9
2014 Feb 17	Eclipse	3.6	12.6 ± 2.0	0.1 ± 2.3
		Period (day)	$3.5247361 \pm 6 \times 10^{-7}$	$3.524750 \pm 1 \times 10^{-6}$
		T_0 (BJD$_{\mathrm{TDB}}$)	$2454560.80567 \pm 8 \times 10^{-5}$	$2454560.80588 \pm 8 \times 10^{-5}$

2005 lightcurves; 65.184 sec for the 2007–2008 lightcurves; 66.184 sec for the 2010–2011 lightcurves; and 67.184 sec for the 2014 lightcurves. We report the results in Table 5.

When all of the mid-times are included in the fit, we obtain $T_0 = 2454560.80567 \pm 8 \times 10^{-5}$ BJD$_{\mathrm{TDB}}$ and $P = 3.5247361 \pm 6 \times 10^{-7}$ day, with a reduced $\chi^2 = 7.0$ indicating a poor overall fit to the data. However, we found that when we exclude the 2010 Jan 19 transit mid-time, we obtain $T_0 = 2454560.80588 \pm 8 \times 10^{-5}$ BJD$_{\mathrm{TDB}}$ and $P = 3.524750 \pm 1 \times 10^{-6}$ day, with a much improved reduced $\chi^2 = 1.0$. It is not clear why our 2010 Jan 19 transit fails to fit a linear ephemeris, although we note that our measured mid-time for this lightcurve is in excellent agreement at the 0.01σ level with the value obtained by Zellem et al. (2014). A full investigation into the cause of this discrepancy is beyond the scope of the current paper. However, given that the other eighteen mid-times are well fit by a linear ephemeris, we conclude that the data is consistent with a constant orbital period.

8 DISCUSSION

The results outlined in Section 7 demonstrate the effectiveness of the GP modelling approach for handling systematics in IRAC lightcurves. Compared with those previously published in the literature, the planet properties inferred from the GP analyses are overall more consistent across different epochs and, for the wavelength-independent properties, across the different wavelength channels. In many cases, this is due to the GP analyses giving uncertainty estimates that are up to ~ 4 times larger than those reported by other authors. For the reasons given in Section 5, we argue that the GP uncertainties provide a more realistic reflection of our ignorance. This is primarily because the GP models offer greater flexibility for handling systematics that do not have a well-understood functional form, compared with the simple parametric approximations used widely in the literature. Marginalisation of the GP model posterior distributions therefore allows us to more exhaustively explore possible degeneracies between the planet signal and systematics, and incorporate these into the uncertainties associated with the inferred planet properties. However, recall from Section 6 that we fixed the covariance parameters to their MLE values before marginalising over the planet parameters with MCMC, which means the uncertainties quoted in Table 4 should in fact be regarded as lower limits to the true uncertainties.

Further verification of the reliability of GP models for inferring planet parameters from IRAC lightcurves could be obtained by systematically applying the method to synthetic datasets, similar to what was done by Gibson (2014). This would allow us to directly compare inferred values for the planet parameters with their known values. The challenge

with such an approach, of course, is that it requires a realistic simulation of the IRAC systematics, for which we do not possess a functional form. With this caveat in mind, however, such an investigation would be useful in the future.

8.1   Atmosphere implications

The IRAC data analysed in the current study address two fundamental hypotheses concerning the nature of HD 209458b's atmosphere. The first of these is the claimed detection of water absorption in transmission made by Beaulieu et al. (2010), based on the larger values for $R_p/R_\star$ that those authors measured for the 5.8μm and 8.0μm channels relative to the 3.6μm and 4.5μm channels. The second is the inference of a thermal inversion in the atmosphere, based on the measurement of deeper eclipses in the 4.5μm and 5.8μm channels relative to the 3.6μm channel by Knutson et al. (2008).

Figure 5 shows the transmission and emission measurements made in the current study with overplotted model spectra computed using the 1D radiative transfer code ATMO (Amundsen et al. 2014; Tremblin et al. 2015), assuming solar abundances and radiative-convective equilibrium. Models were generated under two different scenarios for the day-to-night heat redistribution efficiency: namely, uniform redistribution and zero redistribution (respectively, $\omega = 0.5$ and $\omega = 1$ in Figure 5). Number fractions as a function of atmospheric pressure are given for the major molecules in the lower left panel of Figure 5 and the corresponding pressure-temperature (PT) profiles are shown in the lower right panel of Figure 5.

Our transmission results are in poor agreement with the clear-atmosphere model predictions. In particular, we do not see any enhancement in the opacity at 4.5μm relative to 3.6μm due to water and carbon monoxide absorption. As was mentioned in Section 7.1, our transmission results are also in conflict with those originally presented by Beaulieu et al. (2010). Specifically, Beaulieu et al. measured larger effective radii for the planet in the 5.8 μm and 8.0 μm channels, which coincide with a water absorption band, relative to the 3.6 μm and 4.5 μm channels. We, however, obtain significantly larger uncertainties in the 5.8μm channel and find that the effective radius is constant, or possibly decreasing modestly, across the 3.6–8.0μm IRAC wavelength range. The difference between the results obtained in the current study and those of Beaulieu et al. are likely due to the two approaches used for treating the ramp systematics. For the 5.8 μm channel, Beaulieu et al. truncated the first section of the lightcurve and fit a linear trend in time to the remainder, and for the 8.0 μm channel they decorrelated the ramp using a quadratic polynomial in logarithmic time. The GP model adopted in the current study should be capable of replicating both these explicit functional forms, and indeed, allows marginalisation over an even broader function space (Section 5). We also confirmed that consistent results were obtained with the GP model when sections of varying duration were truncated from the start of the lightcurve. Based on our revised estimates of the planet's effective radii, we conclude that there is no evidence for water absorption in transmission over the IRAC bandpasses. Indeed, Deming et al. (2013) have measured a muted water feature at 1.4μm using WFC3, and interpreted this as evidence for haze in the atmosphere

of HD 209458b. Our results do not contradict this picture, and could suggest that the effect of the opacity inferred by Deming et al. at 1.4μm remains significant out to the IRAC wavelengths.

We also modified the clear-atmosphere transmission model by fitting for an opaque cloud deck to simulate a grey opacity source across the IRAC wavelength range. This was done by allowing the clear atmosphere model to shift vertically while simultaneously setting the absorption to be constant at the cloud deck altitude. Thus our simple model had two tunable parameters: the overall vertical shift of the clear-atmosphere model and the cloud deck altitude. Our best-fit to the data gives a reduced χ^2 of 1.4, with an opaque cloud deck at $R_p/R_\star = 0.1210 \pm 0.003$. As can be seen in the top left panel of Figure 5, the resulting "cloudy" model is simply a horiztonal line, implying that the IRAC data show no evidence for even reduced-amplitude absorption features extending above a cloud deck. If we instead fit a simple flat-line model to the data, where the only free parameter is the vertical level of the opaque cloud deck, the number of degrees of freedom increases from six to seven, and the reduced χ^2 improves to 1.2. Due to the size of our uncertainties, however, we cannot rule out muted absorption features of similar amplitude to the H_2O feature measured by Deming et al. (2013) at 1.4μm with WFC3. Furthermore, we note that our result is consistent at the $\sim 1\sigma$ level with the cloud deck altitude implied by the WFC3 transmission spectrum. Such a comparison should be treated with caution, however, as Deming et al. fixed $a/R_\star = 8.95$ and $i = 86.93°$ in their lightcurve fits (taken from Knutson et al. 2007b), which would introduce an offset in the absolute level those authors derive for $R_p/R_\star$ relative to our study.

For the emission, we produced models both with and without inverted pressure-temperature profiles (lower right panel of Figure 5). To produce thermal inversions in the former models, we artificially included titanium oxide (TiO) and vanadium oxide (VO) with a constant abundance throughout the atmosphere (lower left panel of Figure 5). Our emission results reinforce those of Diamond-Lowe et al. (2014), who argued that there is no evidence for a thermal inversion in the atmosphere of HD 209458b based on the revised estimates for the 4.5μm and 5.8μm eclipse depths. This can be seen in Figure 5, where the models with a thermal inversion are shown to provide a very poor match to the data. For the models without a thermal inversion, the match is better, but still poor. In particular, the latter models underpredict the emission in the 4.5μm channel, where an absorption feature due to carbon monoxide is expected to block radiation emitted from the planetary atmosphere. The fact that we do not detect this absorption is in line with the recent non-detection of carbon monoxide in the dayside hemisphere made by Schwarz et al. (2015) using high-resolution spectroscopy.

We also fit the emission data with a model that assumes the planet radiates as an isothermal blackbody. To generate this model, an ATLAS spectrum (Kurucz 1979, 1993) computed specifically for HD 209458 was used for the stellar emission[3] and the radius ratio was fixed to $R_p/R_\star = 0.121$. Assuming a Planck spectrum for the planet, a temperature

---

[3] http://kurucz.harvard.edu/stars/hd209458
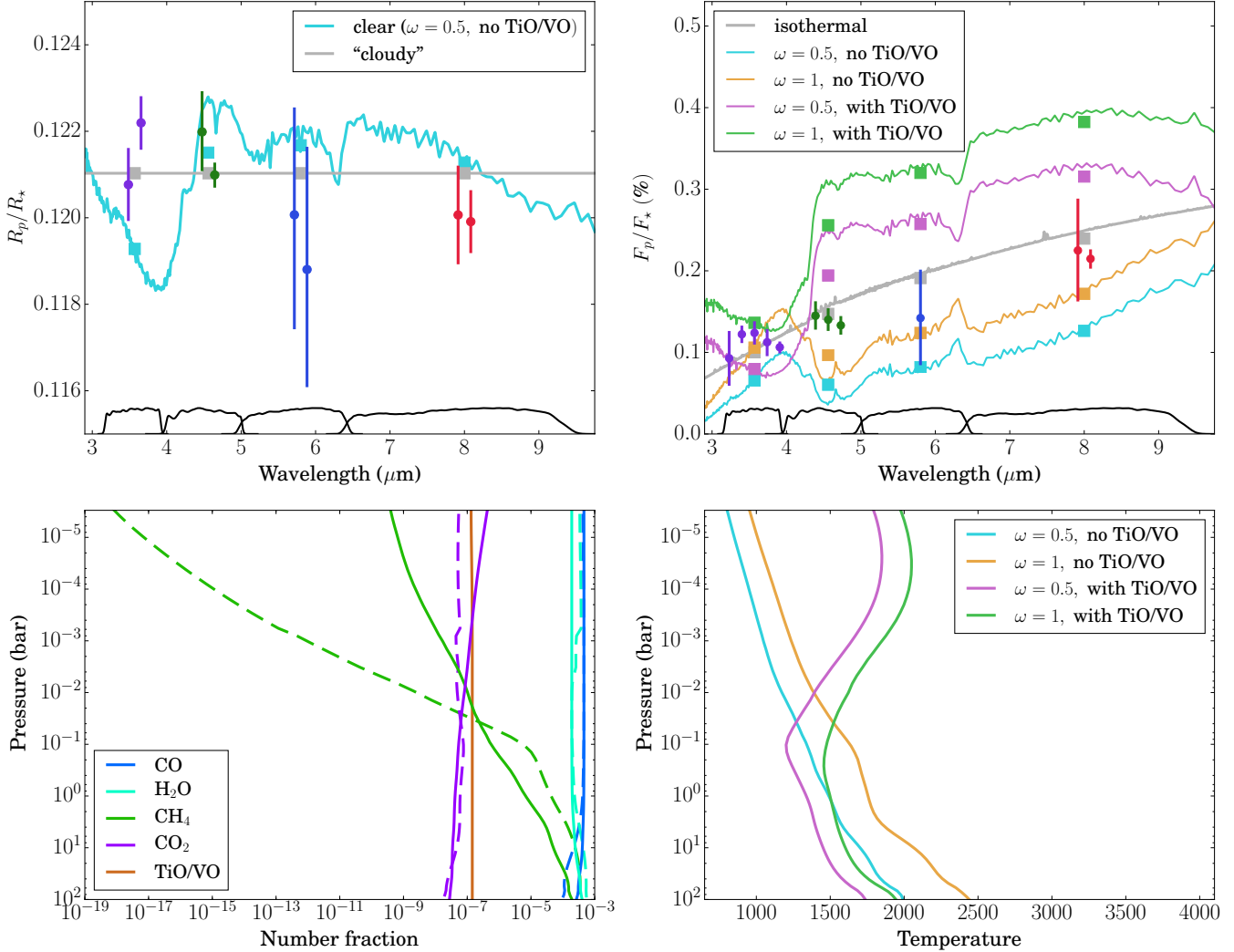
Figure 5. Top panels show the transmission $R_p/R_\star$ (left) and emission $F_p/F_\star$ (right) measurements obtained in the current study for HD 209458b, with colours indicating wavelength channels. IRAC bandpasses are indicated at the bottom of both axes as solid black lines. For measurements obtained in the same wavelength channel, small horizontal offsets have been applied for clarity. Both panels show 1D `ATMO` models assuming solar abundances and chemical equilibrium. Corresponding chemical abundances are shown in the lower left panel and pressure-temperature profiles are shown in the lower right panel. For the abundances, solid lines correspond to uniform heat redistribution from the dayside to the nightside (i.e. $\omega = 0.5$), while dashed lines correspond to zero heat redistribution (i.e. $\omega = 1$). For the transmission, the light blue line shows the clear atmosphere obtained assuming $\omega = 0.5$, while the grey line shows a simple model corresponding to an opaque cloud deck at $R_p/R_\star = 0.1210$, based on fitting a simple cloud model to the data as described in Section 8.1. For the emission, clear atmosphere models are shown for $\omega = 0.5$ and $\omega = 1$, both with and without TiO/VO in the atmosphere. For those models including TiO/VO, the species were assumed to be mixed uniformly throughout the atmosphere, as shown by the brown line in the lower left panel. The inclusion of TiO/VO results in inverted temperature-pressure profiles, as shown in the lower right panel. Also shown in the upper right panel is the emission spectrum obtained using an ATLAS stellar model for HD 209458 and assuming the planet radiates as a blackbody with a temperature of 1484 K. In the top panels, square symbols give the model values integrated over the corresponding IRAC bandpasses.

of 1484 ± 18 K was found to give the best fit to the data, with a reduced χ^2 of 1.5. Indeed, Hansen et al. (2014) have recently put forward the case that blackbody radiation can explain the majority of IRAC emission data that has been published for exoplanets to date, largely due to underestimated uncertainties for the eclipse depths. The acceptable fit provided by the isothermal blackbody for our data supports this hypothesis in the case of HD 209458b, especially consid-

ering the reduced χ^2 could be even lower if we marginalised over the covariance parameters and obtained larger uncertainties (Section 6).

One possibility is that the emission measurements are probing an isothermal layer of the atmosphere above a cloud deck that extends across the dayside hemisphere. This could simultaneously explain the lack of absorption features detected in the IRAC transmission data, and the muted H_2O

feature detected in the WFC3 transmission data, discussed above. We stress, however, that this is a speculative scenario, poorly constrained by the existing observations.

9 CONCLUSION

We have presented an analysis of IRAC transit and eclipse lightcurves for HD 209458b. By binning the lightcurves in time, it was possible to perform the lightcurve analyses using GP models. GPs allow transit lightcurve models to be elegantly defined within a rigorous Bayesian framework. They provide a natural mechanism for handling poorly-understood systematics in the data that are unrelated to the astrophysical signal of interest. Uncertainty is propagated through all levels of the model in a clear and transparent manner, and Occam's razor is automatically implemented, mitigating against overfitting.

We have made a number of significant revisions to previously published results, and in many cases the uncertainties for inferred planet parameters have been increased by factors of $\sim$1–4. This can largely be attributed to the flexibility of the GP models, which allow complex correlations to be handled with a small number of free parameters. The latter point is important, as it means that marginalisation over the model parameter space remains computationally tractable, allowing uncertainties that realistically quantify the degeneracies between the planet signal and instrumental systematics to be derived.

We obtain an overall improvement in consistency for the normalised semimajor axes $a/R_\star$ and orbital inclinations i across different epochs and wavelength channels, compared to results that have been previously published for these datasets. This provides evidence that the GP models are effectively accounting for the systematics, and typically less prone to underestimating uncertainties compared with other lightcurve fitting approaches used in the literature.

The revised GP analyses presented here draw into question a number of claims that have previously been made regarding the atmosphere of HD 209458b, including the detection of water absorption in transmission and the inference of an inverted PT profile for the dayside hemisphere. Instead, our transmission measurements are consistent with a featureless spectrum, and our emission measurements are fit reasonably well assuming the planet radiates as an isothermal blackbody with a temperature of 1484 ± 18 K.

Taken together, our results illustrate how sensitive IRAC analyses are to the systematics treatment. GP analyses have been shown to produce results that are generally more stable, and with uncertainties that are relatively conservative, compared to those obtained using other common approaches. However, we do tend to find good agreement with results obtained using pixel-mapping techniques in the 3.6μm and 4.5μm channels, which is unsurprising given that GPs are in essence quite similar to pixel-mapping. Nonetheless, many results have been published using simpler xy polynomial decorrelations and parametric ramp models, which may not always be adequate. Combined with the lack of spectral resolution afforded by the broad bandpasses, our results suggest that statements made previously in the literature about exoplanet atmospheres relying heavily on the interpretation of IRAC data should be regarded with caution.

REFERENCES

Agol, E., Cowan, N. B., Knutson, H. A., Deming, D., Steffen, J. H., Henry, G. W., & Charbonneau, D. 2010, ApJ, 721, 1861

Amundsen, D. S., Baraffe, I., Tremblin, P., Manners, J., Hayek, W., Mayne, N. J., & Acreman, D. M. 2014, A&A, 564, A59

Ballard, S., Charbonneau, D., Deming, D., Knutson, H. A., Christiansen, J. L., Holman, M. J., Fabrycky, D., Seager, S., & A'Hearn, M. F. 2010, PASP, 122, 1341

Beaulieu, J. P., Kipping, D. M., Batista, V., Tinetti, G., Ribas, I., Carey, S., Noriega-Crespo, J. A., Griffith, C. A., Campanella, G., Dong, S., Tennyson, J., Barber, R. J., Deroo, P., Fossey, S. J., Liang, D., Swain, M. R., Yung, Y., & Allard, N. 2010, MNRAS, 409, 963

Charbonneau, D., Allen, L. E., Megeath, S. T., Torres, G., Alonso, R., Brown, T. M., Gilliland, R. L., Latham, D. W., Mandushev, G., O'Donovan, F. T., & Sozzetti, A. 2005, ApJ, 626, 523

Charbonneau, D., Knutson, H. A., Barman, T., Allen, L. E., Mayor, M., Megeath, S. T., Queloz, D., & Udry, S. 2008, ApJ, 686, 1341

Claret, A. 2004, A&A, 428, 1001

Crossfield, I. J. M., Knutson, H., Fortney, J., Showman, A. P., Cowan, N. B., & Deming, D. 2012, ApJ, 752, 81

Deming, D., Harrington, J., Laughlin, G., Seager, S., Navarro, S. B., Bowman, W. C., & Horning, K. 2007, ApJ, 667, L199

Deming, D., Harrington, J., Seager, S., & Richardson, L. J. 2006, ApJ, 644, 560

Deming, D., Knutson, H., Kammer, J., Fulton, B. J., Ingalls, J., Carey, S., Burrows, A., Fortney, J. J., Todorov, K., Agol, E., Cowan, N., Desert, J.-M., Fraine, J., Langton, J., Morley, C., & Showman, A. P. 2014, ArXiv e-prints

Deming, D., Wilkins, A., McCullough, P., Burrows, A., Fortney, J. J., Agol, E., Dobbs-Dixon, I., Madhusudhan, N., Crouzet, N., Desert, J.-M., Gilliland, R. L., Haynes, K., Knutson, H. A., Line, M., Magic, Z., Mandell, A. M., Ranjan, S., Charbonneau, D., Clampin, M., Seager, S., & Showman, A. P. 2013, ApJ, 774, 95

Désert, J.-M., Lecavelier des Etangs, A., Hébrard, G., Sing, D. K., Ehrenreich, D., Ferlet, R., & Vidal-Madjar, A. 2009, ApJ, 699, 478

Diamond-Lowe, H., Stevenson, K. B., Bean, J. L., Line, M. R., & Fortney, J. J. 2014, ApJ, 796, 66

Eastman, J., Siverd, R., & Gaudi, B. S. 2010, PASP, 122, 935

Evans, T. M., Pont, F., Sing, D. K., Aigrain, S., Barstow, J. K., Désert, J.-M., Gibson, N., Heng, K., Knutson, H. A., & Lecavelier des Etangs, A. 2013, ApJ, 772, L16

Gelman, A. & Rubin, D. B. 1992, Stat. Sci., 7, 457

Gibson, N. P. 2014, MNRAS, 445, 3401

Gibson, N. P., Aigrain, S., Barstow, J. K., Evans, T. M., Fletcher, L. N., & Irwin, P. G. J. 2013a, MNRAS, 428, 3680

—. 2013b, MNRAS, 436, 2974

Gibson, N. P., Aigrain, S., Pont, F., Sing, D. K., Désert, J.-M., Evans, T. M., Henry, G., Husnoo, N., & Knutson, H. 2012a, MNRAS, 422, 753

Gibson, N. P., Aigrain, S., Roberts, S., Evans, T. M., Osborne, M., & Pont, F. 2012b, MNRAS, 419, 2683

Goldberg, P. W., Williams, C. K. I., & Bishop, C. M. 1998, in In Advances in Neural Information Processing Systems 10 (MIT Press), 493–499

Hansen, C. J., Schwartz, J. C., & Cowan, N. B. 2014, MNRAS, 444, 3632

Hastings, W. K. 1970, Biometrika, 57, 97

Hayek, W., Sing, D., Pont, F., & Asplund, M. 2012, A&A, 539, A102

Knutson, H. A., Charbonneau, D., Allen, L. E., Burrows, A., & Megeath, S. T. 2008, ApJ, 673, 526

Knutson, H. A., Charbonneau, D., Allen, L. E., Fortney, J. J., Agol, E., Cowan, N. B., Showman, A. P., Cooper, C. S., & Megeath, S. T. 2007a, Nature, 447, 183

Knutson, H. A., Charbonneau, D., Noyes, R. W., Brown, T. M., & Gilliland, R. L. 2007b, ApJ, 655, 564

Knutson, H. A., Lewis, N., Fortney, J. J., Burrows, A., Showman, A. P., & et al. 2012, ApJ, 754, 22

Kurucz, R. 1993, ATLAS9 Stellar Atmosphere Programs and 2 km/s grid. Kurucz CD-ROM No. 13. Cambridge, Mass.: Smithsonian Astrophysical Observatory, 1993., 13

Kurucz, R. L. 1979, ApJS, 40, 1

Lewis, N. K., Knutson, H. A., Showman, A. P., Cowan, N. B., Laughlin, G., Burrows, A., Deming, D., Crepp, J. R., Mighell, K. J., Agol, E., Bakos, G. Á., Charbonneau, D., Désert, J.-M., Fischer, D. A., Fortney, J. J., Hartman, J. D., Hinkley, S., Howard, A. W., Johnson, J. A., Kao, M., Langton, J., & Marcy, G. W. 2013, ApJ, 766, 95

Machalek, P., Greene, T., McCullough, P. R., Burrows, A., Burke, C. J., Hora, J. L., Johns-Krull, C. M., & Deming, D. L. 2010, ApJ, 711, 111

Mandel, K. & Agol, E. 2002, ApJ, 580, L171

Mchutchon, A. & Rasmussen, C. E. 2011, in Advances in Neural Information Processing Systems 24, ed. J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Weinberger (Curran Associates, Inc.), 1341–1349

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., & Teller, A. H. 1953, J. Chem. Phys., 21, 1087

Muller, R. A. & Buffington, A. 1974, Journal of the Optical Society of America (1917-1983), 64, 1200

Nelder, J. A. & Mead, R. 1965, Computer Journal, 7, 308

Pont, F., Husnoo, N., Mazeh, T., & Fabrycky, D. 2011, MNRAS, 414, 1278

Rasmussen, C. E. & Williams, C. 2006, Gaussian Processes for Machine Learning (Cambridge, Massachusetts, USA: MIT Press)

Schwarz, G. E. 1978, Ann. Stat., 6, 461

Schwarz, H., Brogi, M., de Kok, R., Birkby, J., & Snellen, I. 2015, A&A, 576, A111

Seager, S. & Deming, D. 2010, ARA&A, 48, 631

Stevenson, K. B., Harrington, J., Fortney, J. J., Loredo, T. J., Hardy, R. A., Nymeyer, S., Bowman, W. C., Cubillos, P., Bowman, M. O., & Hardin, M. 2012, ApJ, 754, 136

Stevenson, K. B., Harrington, J., Nymeyer, S., Madhusudhan, N., Seager, S., Bowman, W. C., Hardy, R. A., Deming, D., Rauscher, E., & Lust, N. B. 2010, Nature, 464, 1161

Todorov, K. O., Deming, D., Burrows, A., & Grillmair, C. J. 2014, ApJ, 796, 100

Torres, G., Winn, J. N., & Holman, M. J. 2008, ApJ, 677, 1324

Tremblin, P., Amundsen, D. S., Mourier, P., Baraffe, I., Chabrier, G., Drummond, B., Homeier, D., & Venot, O. 2015, ArXiv e-prints, accepted by ApJL

Zellem, R. T., Lewis, N. K., Knutson, H. A., Griffith, C. A., Showman, A. P., Fortney, J. J., Cowan, N. B., Agol, E., Burrows, A., Charbonneau, D., Deming, D., Laughlin, G., & Langton, J. 2014, ApJ, 790, 53

Table A1. Maximum likelihood estimates for the GP covariance parameters. These are the values that were fixed for the MCMC analyses described in Section 6.

Channel (μm)	Date	Signal	A_{xy} (%)	L_x (pix)	L_y (pix)	A_t (%)	L_t (min)	σ_w (ppm)
3.6	2005 Nov 28	Eclipse	2.0552	0.642	0.528	0.0444	37.117	1541
	2007 Dec 31	Transit	0.4874	0.169	0.187	0.0216	20.708	2185
	2008 Jul 19	Transit	0.3089	0.138	0.136	0.0677	0.319	2066
	2011 Jan 12	Eclipse (1st)	1.0988	0.176	0.091	0.0261	5.471	499
	2011 Jan 14	Transit	0.6342	0.127	0.145	0.1754	0.322	344
	2011 Jan 16	Eclipse (2nd)	1.8569	0.235	0.233	0.0331	6.710	541
	2014 Feb 13	Eclipse (1st)	1.5366	0.167	0.135	0.0398	5.574	553
	2014 Feb 15	Transit	1.6826	0.195	0.196	0.0488	0.553	660
	2014 Feb 17	Eclipse (2nd)	1.2590	0.151	0.118	0.0234	3.089	453
4.5	2005 Nov 28	Eclipse	0.8656	0.238	0.254	0.8796	2386.526	1391
	2008 Jul 22	Transit	0.0001	143.003	109.329	0.0302	5.887	3148
	2010 Jan 18	Eclipse (1st)	0.2467	0.067	0.090	0.0211	25.212	578
	2010 Jan 19	Transit	0.8769	0.255	0.231	0.0086	24.042	569
	2010 Jan 21	Eclipse (2nd)	0.8468	0.235	0.270	0.0182	15.931	566

Channel (μm)	Date	Signal	A_τ (%)	L_τ (log[min])	h (min)	A_t (%)	L_t (min)	σ_w (ppm)
5.8	2005 Nov 28	Eclipse	0.0732	6.187	2796.4	0.0575	195.8	3956
	2007 Dec 31	Transit	0.3080	6.127	489.8	0.0000	67.9	3037
	2008 Jul 19	Transit	0.7993	12.986	0.1	0.0000	$> 10^4$	2986
8.0	2005 Nov 28	Eclipse	0.2781	6.277	1142.0	0.0000	$> 10^4$	3749
	2007 Dec 24	Transit	0.0000	357.352	291853.4	0.2427	1221.2	1163
	2007 Dec 25	Eclipse	0.0163	5.887	12729.4	0.0000	$> 10^4$	1215
	2008 Jul 22	Transit	1.8164	9.791	1.0	0.0000	$> 10^4$	2213

APPENDIX A: COVARIANCE PARAMETER MAXIMUM LIKELIHOOD ESTIMATES

Table A1 reports the maximum likelihood estimates for the GP covariance parameters that were fixed for the MCMC analyses, as described in Section 6.

This paper has been typeset from a TEX/ LATEX file prepared by the author.